1968

Further test of the Yerkes-Dodson law

Donald Arthur Maguire

The University of Montana

1048

A FURTHER TEST OF THE YERKES-DODSON LAW

By

Donald A. Maguire

B. A. Northern Arizona University, 1966

Presented in partial fulfillment of the requirements for the degree of

Master of Arts

UNIVERSITY OF MONTANA

1968

Approved by:

Harold Babb

Chairman, Board of Examiners

Fred S Honkala

Dean, Graduate School

June 4, 1968

Date

6-10-68

2449-8T

ACKNOWLEDGMENTS

The author would like to express his appreciation to the members of his thesis committee, Dr. Harold Babb, Chairman; Dr. James M. Whitehouse; and Dr. Richard M. Boehmler, for their criticism and consideration, and to Mr. Thomas Danielson for his assistance in constructing the apparatus.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

The Yerkes-Dodson law, a generalization relating the variables of task difficulty and motivation level during acquisition of a learning task, has historically found wide acceptance among psychologists. It has been proclaimed "confirmed" by several writers (Broadhurst, 1957; Hall, 1961; Young, 1936) and it has been invoked in the interpretation of the findings of numerous investigations. In making reference to the law, however, few seem to have noticed that it exists in many forms (see Appendix A). Several interpretations, explanations, and reformulations still impart to it a generality that goes far beyond the original statement of the law (Yerkes, 1909; Yerkes & Dodson, 1908). Furthermore, Brown (1965) has reviewed the evidence in support of the law and found it to be inadequate. Consequently, the status and applicability of the Yerkes-Dodson law are presently in doubt.

In their early experiment, Yerkes and Dodson (1908) trained dancing mice on a black-white avoidance discrimination task with three levels of task difficulty and several levels of shock motivation. It was found that the most rapid learning on the difficult task was exhibited at low levels of shock while the most rapid learning on the easy- and medium-difficulty tasks was exhibited at higher levels of shock. In their conclusions the experimenters stated: "As the difficultness of discrimination is increased the strength of that stimulus which is most favorable to habit-formation approaches the threshold (p.481)." It was this generalization which eventually became known as the Yerkes-Dodson law.

A similar experiment by Cole (1911), with chickens as subjects (Ss), produced results which were interpreted as supporting the law. Dodson (1915) attempted to extend the generalization to kittens and obtained results which were consistent with the findings of the original experiment. Much later, Broadhurst (1957) used an underwater maze with rats as Ss and found a significant interaction between motivation (degree of air deprivation) and task difficulty (brightness ratio between pairs of discriminanda). Broadhurst interpreted his results as confirming the law, after reformulating it to state: "...the optimum motivation for a learning task decreases with increasing difficulty (p.345)."

The foregoing experiments, which provide the primary evidence for the law, have met with serious criticism (see Appendix B for further description and related studies). Broadhurst (1957; 1959) pointed out the difficulty of accurately controlling shock levels in the early studies by Cole (1911), Dodson (1915), and Yerkes and Dodson (1908), and suggested that too few subjects were employed in these studies. Brown (1965) has pointed out shortcomings in all of the studies designed to test the law, including the study by Broadhurst (1957). The criticisms registered by Brown are listed below, along with the studies to which each pertains:

1. Equivocal interpretation of results (Broadhurst, 1957; Cole, 1911; Dodson, 1915; Yerkes & Dodson, 1908).

2. Lack of a factorial design (Dodson, 1915; Yerkes & Dodson, 1908).

3. Weak definition of "optimum level" (Broadhurst, 1957; Cole, 1911; Yerkes & Dodson, 1908).

It should also be noted that none of these experiments employed tasks which were previously established as being of differential difficulty, and in the early experiments by Cole (1911), Dodson (1915), and Yerkes and Dodson (1908) there was no real assessment of the actual effectiveness of the manipulation of task difficulty.

Brown (1965) has further indicated that all of these studies have failed to meet the requirements of a proof of the law. According to Brown, a conclusive study must: 1) demonstrate that an optimum level of motivation can be found for each level of task difficulty, and 2) demonstrate an interaction between motivation and task difficulty, with the optimum levels of motivation ordered along a continuum from lowest for difficult tasks to highest for easy tasks. For example, with three levels of difficulty the optimum levels of motivation should be ordered Difficult < Medium < Easy. Although none of the previous experiments have provided conclusive evidence for the law according to Brown's criteria, the results from those studies are consistent with the original formulation of the law by Yerkes (1909) and Yerkes and Dodson (1908) and indicate that the suggested relation remains a strong possibility.

The present experiment was designed to provide a rigorous test of the principle suggested by Yerkes and Dodson (1908) as interpreted by the present experimenter. A factorial design was employed having five levels of food deprivation and three levels of discrimination task difficulty, with the experimental conditions arranged so as to avoid the weaknesses enumerated by Broadhurst (1957; 1959) and Brown (1965). Food deprivation was chosen as the drive operation in this experiment for two main reasons. First, several writers have suggested that the law be tested with some form of appetitive motivation (Broadhurst, 1957, 1959; Brown, J., 1961; Brown, W. P., 1965; Cofer & Appley, 1964). Second, studies investigating the Yerkes-Dodson law have all employed some form of aversive motivating condition and, according to Brown's criteria, all have failed to provide convincing evidence for the law, suggesting that aversive conditions may in some way limit the outcome of any experiment designed to demonstrate the law. It is generally recognized that aversive conditions, such as shock- and water-escape, differ from appetitive conditions, such as food and water deprivation, in several respects, including the rapidity of onset of the motivating condition, specificity of the condition to a particular situation, and the possibility that the drive is mediated by fear. If, as suggested, one of these characteristics of aversive conditions limits the response possibilities in a Yerkes-Dodson type of situation, the use of an appetitive condition, food deprivation, should alleviate this problem to a great extent and facilitate a clear demonstration of the law.

One of the major shortcomings of previous studies has been the use of inadequately defined terms. "Level of motivation" is here interpreted as "drive level", defined in terms of hours of food deprivation. "Optimum level of motivation" is interpreted as "optimum drive level", defined as "that level of drive at which the most rapid learning occurs on a task, provided that such learning (as reflected in learning scores) is significantly faster than learning at one or more of the other drive levels on that task." In order to avoid confusion about the statement of the law to be tested, the following formulation was constructed for purposes of this study: "The optimum drive level for the acquisition of a discrimination task is inversely related to the difficulty of the task." This statement includes all of the essential aspects of the law and is consistent with the original version (Yerkes, 1909; Yerkes & Dodson, 1908).

II. METHOD

Subjects

The Ss were 120 male Sprague-Dawley albino rats obtained from Northwest Rodent Supply, Pullman, Washington. They were approximately 90 days of age on the first day of handling.

Apparatus

The apparatus was a discrimination box (see Figure 1) constructed of wood and painted flat black with a 5-in. high by 5-in. wide starting area extending for 12 in. and then gradually widening in a straight line over a distance of 6 in. to a width of 10 1/4 in. Adjoining the wide end of the starting area were 2 parallel discrimination alleys 5 in. wide by 5 in. high by 14 1/2 in. long, separated by a 1/4-in. thick plywood panel 5 in. high extending the length of the alleys. At the distant end of each alley a metal food cup 3 in. in diameter was recessed into the floor so that the top of the cup was flush with the floor level. Behind each food cup was mounted a 5-in. high by 4 5/8-in. wide frosted glass screen which faced the alley. Mounted 3/8 in. behind each of these was a second screen of the same material and dimensions. Extending beyond each of these screens was another alley 5 in. high by 4 5/8 in. wide which allowed for passage of light from a projector, one arm of which was centered spatially in each passageway 24 1/2 in. behind the second screen. The two passageways were separated by a 5-in. high panel and covered with a 1/4 in.-thick plywood lid which made them light-tight.

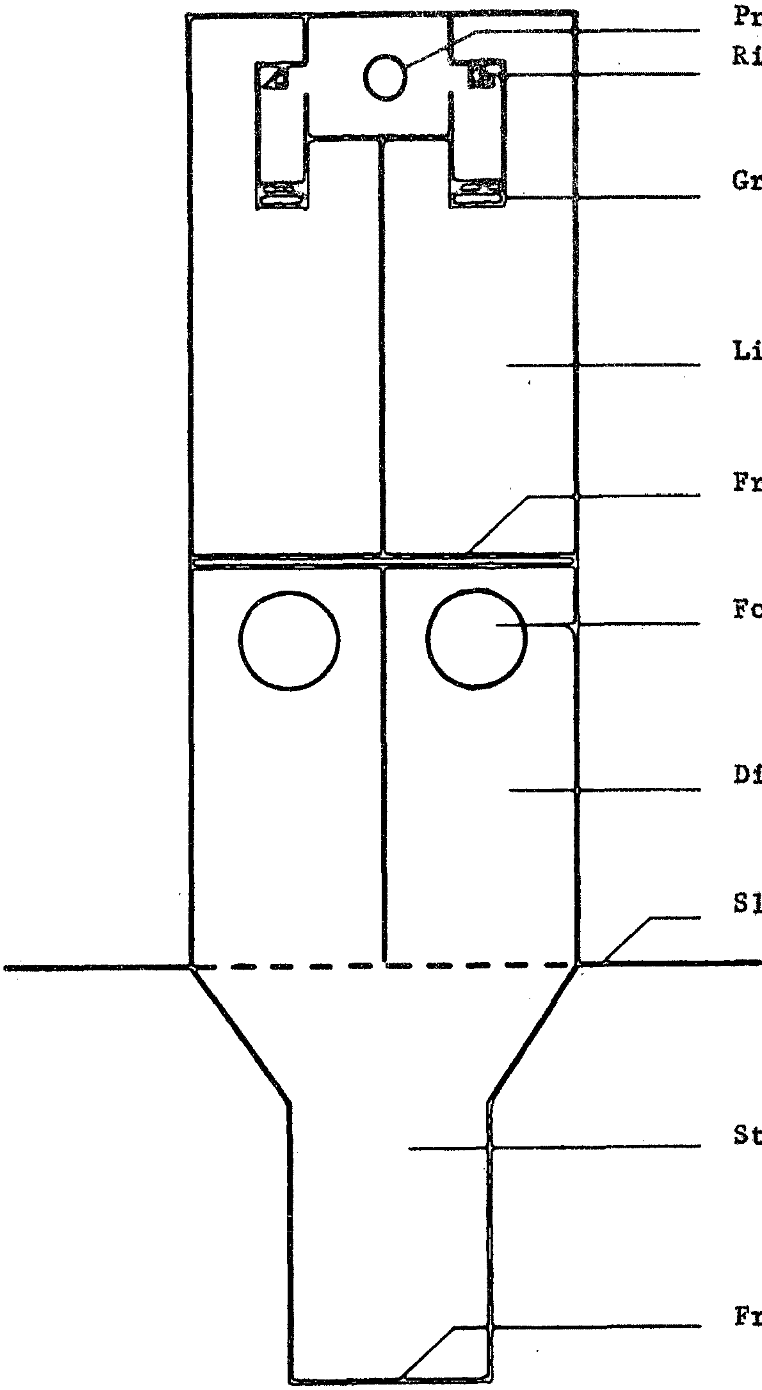

Figure 1. Diagramatical top view of discrimination apparatus.

The projector consisted of a 6-in. long by 3-in. wide by 5 1/2-in. high wooden enclosure which allowed light to pass out through openings 2 in. high by 1 in. wide cut into the center of each side. Each opening was fitted with a right-angle prism which directed light through a short arm and down a passageway to the frosted glass screens. The arms of the projector were 1 3/4 in. wide by 1 7/8 in. high by 5 in. long with grooves at the end distal to the light source, providing for the insertion of Wratten Neutral Density filters which controlled the intensity of light reaching the screens. Located in the center of the projector was a 200-w. Sylvania projection lamp no. CGW/CGT inserted into a pre-focus base. The lamp was cooled by a Union Blower exhaust fan through a passage for air provided at the top and bottom of the projector.

The discrimination alleys were covered by a 1/4-in. plywood lid which was hinged near the frosted glass screens so that animals could be easily removed from the alleys. The starting area was covered by a 1/2-in. thick piece of clear plexiglass and equipped with a front-opening plexiglass door. On each side of the juncture of the starting area and the discrimination alleys was a sliding plexiglass door which could be moved to retain animals in the alley. The only illumination provided for the experimental room was a red 25-w. light bulb located on the ceiling directly above the apparatus.

Design

Five drive levels and three levels of task difficulty were used, providing a 5 x 3 factorial design (15 treatments) with 8 replications (8 Ss per treatment group). The drive levels employed were 4, 10, 22, 36, and 44 hrs. of food deprivation. These levels were chosen in order to cover a wide range of drive while providing an administration schedule which would be reasonably convenient for the experimenter. The light intensities used were determined in a preliminary study in which 5 groups of 10 rats each were trained on 5 combinations of intensities while under 22 hr. food deprivation (see Appendix C). An analysis of the results yielded 3 tasks which produced significantly different mean learning scores. The combinations of Wratten filters which provided these tasks were as follows: Easy - .5 and 3.5 log reduction units (lru); Medium - 2.0 and 3.0 lru; Difficult - 2.0 and 2.5 lru. These combinations constituted the 3 levels of task difficulty used in the present experiment.

Procedure

On the initial day of handling experience the Ss were randomly assigned to 15 experimental groups with 8 Ss in each group. Each S was handled for 2 min. a day in an animal housing room for 10 consecutive days. During the next 5 days each S was handled for 3 min. a day in the experimental room with food available while the S was being handled. Following the second day of handling each group was placed on the deprivation schedule appropriate for that experimental condition.

Each S was given 5 trials per day on alternate days until it reached a criterion of 18 correct out of 20 consecutive trials. A non-correction method was employed and the inter-trial interval was approximately 6 min. The brighter alley was designated as correct and the order of presentation of the positive stimulus followed the

the repeated sequence of LRRLLRLRRLRLLRRLRLLR. If a S responded to the same side on three or more consecutive trials the positive stimulus then appeared in the opposite alley until the animal responded to that side, at which time the above order of presentation was resumed. A 1-in. bar of Purina Laboratory Chow was present in the food cup of the correct alley, and each S was retained for 30 sec. after entering one of the alleys.

III. RESULTS

The number of trials required to reach the criterion of 18 correct out of 20 consecutive trials was recorded for each S. However, after a large number of trials it became apparent that many of the Ss in the 4-hr. group and a few Ss in the 10-hr. group were showing little or not improvement. They were exhibiting strong position responses and taking excessive periods of time to enter one of the alleys. In order to avoid prolonging the experiment indefinitely it was decided that when all animals in the 44-, 36-, and 22-hr. groups had reached criterion, any remaining animals in the 4- and 10-hr. groups would be terminated and the minimum number of trials in which they could have reached criterion would be added to their terminal scores. This procedure was considered to be justified since continuation of the experiment could only serve to increase the significance of differences between these drive groups and the remaining drive groups at each level of task difficulty.

Bartlett's Test for homogeneity of variance was applied to the resulting data, revealing significant heterogeneity of variance (X^2 = 34.08, df = 14, $p < .05$). Such heterogeneity of variance clearly violates the assumptions of parametric analysis. However, writers such as Boneau (1960), Box (1953), and Edwards (1961) have suggested that the F test is relatively insensitive to departures from normality and is little influenced by heterogeneity of variance when treatment groups are of equal size and the overall N is large. Since the N in the present experiment was quite large and the treatment groups were of equal size, it was decided to proceed with the analysis of variance, but to evaluate the data by non-parametric methods as well.[1]

Means and standard deviations of trials to criterion scores for all experimental groups are presented in Table 1, and mean scores for all groups are presented graphically in Figure 2. An analysis of variance of trials to criterion scores (Table 2) revealed a significant Task effect ($F = 29.57$, $df = 2/105$, $p < .005$) and Drive effect ($F = 24.52$, $df = 4/105$, $p < .005$). The Task x Drive interaction term, which is of primary concern, was also significant ($F = 2.196$, $df = 8/105$, $p < .05$), and part of the prediction was thus fulfilled.

Duncan's New Multiple Range Test (Table 3) was used to compare the means from each level of task difficulty. All tests were two-tailed, and a p value of .05 was accepted as significant. The lowest mean for each task was significantly different from at least two other means for that task and thus, according to the preset criterion, there was an optimum drive level for each task. The Mann-Whitney U test (Table 4) also confirmed that the lowest mean from each task differed significantly from at least two other means from that task.

The order of the lowest task means also satisfied the established criterion for optimum drive levels. For the Easy task the optimum drive level was 44 hrs. deprivation, for the Medium task it was 36 hrs., and

1. When the scores of the 4-hr. group were omitted, an analysis of the data revealed no significant heterogeneity of variance and produced results thoroughly consonant with those presented here. That analysis may be found in Appendix D, along with an evaluation of error score and run speed data.

Table 1.

Means and standard deviations of trials to criterion.

		Deprivation Condition				
		4 hr.	10 hr.	22 hr.	36 hr.	44 hr.
Easy Task	M	185.62	208.00	86.12	93.25	72.50
	SD	78.64	69.84	20.84	29.43	23.29
Medium Task	M	205.87	212.68	136.37	131.12	159.50
	SD	60.99	32.80	35.15	28.46	31.81
Difficult Task	M	238.25	221.87	165.75	183.37	193.25
	SD	34.26	33.16	28.46	31.86	33.46

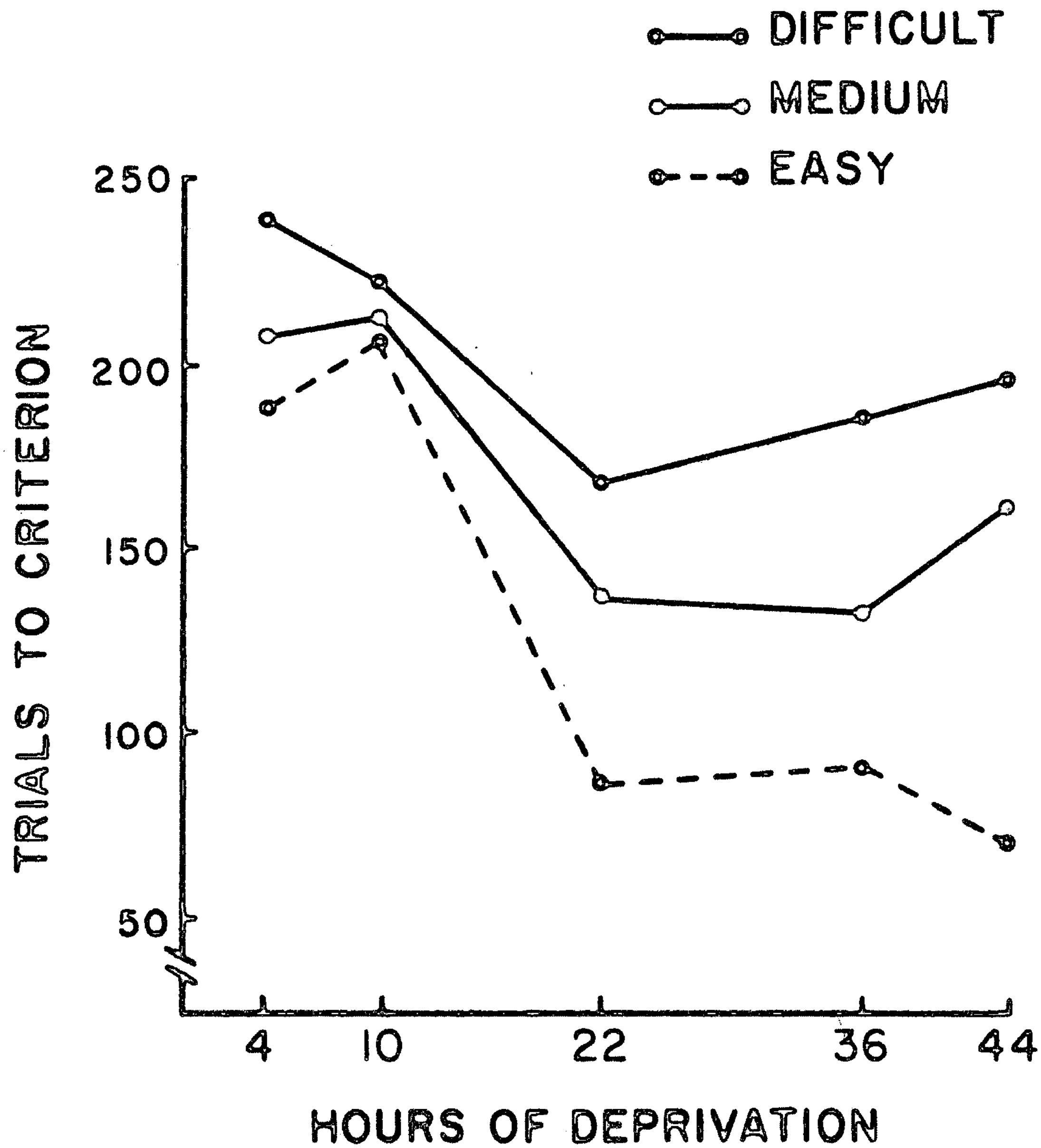

Figure 2. Mean number of trials to criterion by level of task difficulty and drive level. Each point represents the mean trials to criterion for 8 Ss.

Table 2.

Summary of trials to criterion analysis of variance.

Source of Variance	Sum of Squares	df	Mean Square	F
Task (T)	102,452.27	2	51,226.13	29.57*
Drive (D)	169,916.11	4	42,479.02	24.52*
T x D	30,430.05	8	3,803.75	2.196**
Error	181,871.04	105	1,732.10	
Total	484,669.47	119		

*$p < .005$
**$p < .05$

Table 3.

Significant differences between drive groups at each level of task difficulty in trials to criterion (Duncan's New Multiple Range Test).

	Deprivation Condition			
	10 hr.	22 hr.	36 hr.	44 hr.
Easy Task				
4 hr.	-	.05	.05	.05
10 hr.		.05	.05	.05
22 hr.			-	-
36 hr.				-
Medium Task				
4 hr.	-	.05	.05	.05
10 hr.		.05	.05	.05
22 hr.			-	-
36 hr.				-
Difficult Task				
4 hr.	-	.05	.05	.05
10 hr.		.05	.05	-
22 hr.			-	-
36 hr.				-

Table 4.

Significance of differences in trials to criterion between drive group with lowest mean and remaining drive groups at each level of task difficulty (Mann-Whitney U Test).

	Deprivation Condition				
	4 hr.	10 hr.	22 hr.	36 hr.	44 hr.
Drive Group with Lowest Mean					
Easy Task (44 hr., M = 72.50)	.001	.001	.139	.08	-
Medium Task (36 hr., M = 131.12)	.005	.001	.360	-	.041
Difficult Task (22 hr., M = 165.75)	.001	.002	-	.097	.052

for the Difficult task it was 22 hrs. Since there are five possible optimum levels for each task and three different tasks, there are 125 ways in which the experiment as a whole might have turned out. Ten of those possible outcomes fit the pattern of Difficult $<$ Medium $<$ Easy, and the present results represent one of the orders which fit that pattern. The observed outcome, then, has a probability of occurence of 10/125 = .08, and therefore confirms the predicted order of optimum drive levels at the .08 level of significance, the highest confirmation possible with the experimental design used.

IV. DISCUSSION

The present results strongly support the experimenter's formulation of the Yerkes-Dodson law. In all respects the conditions for a demonstration of the law have been met and the law is therefore confirmed for this experimental situation. Whether the law applies to other tasks or to situations in which other drives are operative is not clear at present. While earlier research (Broadhurst, 1957; Cole, 1911; Dodson, 1915; Yerkes & Dodson, 1908) has been criticized as failing to provide strong support for the law, the results of that research are nevertheless consistent with the law and seem to suggest that it may be applicable beyond the present situation. In that respect it is of special interest to note that the present experiment employed a different species and different motivating conditions than those used originally by Yerkes and Dodson (1908), suggesting that difficulty in demonstrating the predicted outcome may be due to problems of experimental design or limitations imposed by the experimental conditions.

Several factors deserve serious consideration in comparing the results of the present study with earlier studies which failed to provide such a clear demonstration of the law. The various experiments investigating this law have not all used the same performance measure and this might account for at least some of the differences in results. Yerkes and Dodson (1908) used a rather stringent learning criterion of 30 consecutive correct trials (3 errorless days) while Broadhurst (1957) measured the number of correct trials out of 100. In the present experiment a criterion of 18 correct out of 20 consecutive trials was employed. It is possible that these measures are sensitive to different aspects of an animal's performance and therefore measure response acquisition differently. Furthermore, there are differences in the response requirements of each experimental situation, and these differences might also affect the speed of learning at various drive levels.

The relative difficulty of the tasks employed in the present experiment as compared with those used by Yerkes and Dodson (1908) and Broadhurst (1957) is not known, but it is possible that the three tasks used here were either more difficult or less difficult or they may have covered a wider range of difficulty. In previous studies the designations of easy, medium, and difficult tasks were made on the basis of brightness ratios determined by a light meter reading or on the basis of apparent difficulty. In the present study, however, the levels of illumination used were chosen only after it was experimentally demonstrated that different numbers of trials were required to learn these tasks when drive level was held constant (see Appendix C). Between-subject variability may easily obscure differences when tasks are not of sufficiently different difficulty, and the procedure of using pre-tested tasks in the present experiment probably separated the tasks enough to overcome the negatively-disposing effects of such variability. It may be the case, then, that the tasks used by earlier investigators did not cover a wide enough range of task difficulty to provide differences which would be statistically significant.

Another factor of importance in comparing the results of these experiments is the type of motivation employed. As previously mentioned, all prior studies directly relating to the Yerkes-Dodson law employed some form of aversive motivation (see also Appendix B). Since even a relatively small amount of aversive stimulation often results in a strong response (e.g., with 0.0 seconds delay Broadhurst obtained faster responding than was observed in the 22- or 36-hr. deprivation condition in the present experiment), the use of aversive conditions may put a lower limit on performance, essentially restricting the response range available for measurement. If speed of acquisition, like response strength, is related to drive strength as stated in the Yerkes-Dodson law, the possiblity of demonstrating an interaction between drive strength and task difficulty would be decreased in any situation employing a motivating condition which only produces relatively strong responding. It may be the case that the law is more easily demonstrated with appetitive motivation. In the opinion of the present writer, however, if both task and drive levels are carefully selected and cover a wide enough range the relationship should be observable using virtually any type of motivating condition which is quantifiable.

Owing to differences in the type of motivation employed, the distribution of drive levels selected for study in each case deserves careful consideration in comparing the results of these different experiments. In the present study a wide range of drive levels was used and statistically significant inter-group performance differences were demonstrated. Whether drive levels were adequately distributed in other experiments cannot be assessed since performance differences between these levels have not been reported in terms of statistical significance. It may be the case that the Yerkes-Dodson law is limited to, or at least can only be demonstrated in, situations in which there are relatively large differences in the drive levels involved. In this respect the definition of optimum level used here is open to criticism since in any case the inclusion of extreme drive levels would most likely result in significant within-task differences. A better definition of the optimum level would probably require that the lowest drive group mean for any task be significantly different from every other drive group mean on that task, but this would necessitate using levels even more widely separated than those in the present study and consequently a poorer estimate of the actual location of such an optimum level would be obtained.

There is an apparent dilemma in this situation, making the optimum level essentially impossible to locate. In order to isolate an optimum level it would be necessary to use closely related levels of drive, but it becomes extremely difficult to demonstrate differences between closely related drive levels. The precise specification of optimum levels thus becomes a practical impossibility. Therefore, it is probably best simply to keep in mind that optimum levels are in every case only approximate, and to consider optimum levels more as descriptions of regions than precise points. Further experimentation in which the parameters of drive level and task difficulty are systematically examined should provide additional information about the nature and location of optimum levels and thereby be of help in formulating a more adequate definition of the term.

Although the present experiment has clearly demonstrated the phenomenon predicted by the Yerkes-Dodson law, further investigation of the conditions under which the law holds is still needed. The parameters of drive level and task difficulty need to be carefully studied in relation to one another under conditions of hunger motivation and other conditions as well, first in brightness discrimination situations and then with other tasks. Parametric investigation of these variables can alone begin to answer the questions raised here of how widely separated tasks must be to demonstrate performance differences in acquisition, whether the minimum differences between tasks necessary to produce differential learning rates changes with the level of the difficulty range involved (i. e., is there a ceiling on difficulty?), and how widely separated drive levels must be at various levels of task difficulty to produce differential learning (i. e., do drive and task difficulty show less interaction at high levels of task difficulty?). The information obtained in such studies would be invaluable in designing subsequent experiments to determine the applicability of the law to other forms of motivation and other types of tasks.

In light of the fact that the law has presently only been tested in a highly limited number of situations it might be more appropriate to refer to it as a "principle" or "phenomenon" rather than as a "law." However, since the term "law" has historically been applied, it will probably be less confusing if the present designation is continued but the limitations herein set forth be recognized. If, as suspected, the law eventually finds support in other experimental situations, such extension would be of considerable theoretical and practical interest, for as Broadhurst (1957, 1959) has already noted, it is a law of potentially great significance to most areas of psychology.

V. SUMMARY

In order to test the Yerkes-Dodson law, rats were trained on a brightness discrimination task in a 5 x 3 factorial experiment having five levels of food deprivation and three levels of discrimination difficulty. An analysis of variance of trials to criterion indicated a significant interaction between task difficulty and level of motivation, and significant differences between drive levels were found at each level of task difficulty. Furthermore, the drive levels producing the fastest learning at each level of task difficulty were ordered Difficult < Medium < Easy, confirming the predictions of the law.

The results presented here provide convincing evidence of the validity of the law, and are of special significance in that the law has not previously been demonstrated with food deprivation as the motivating condition. Incongruities among the studies relating to the law were discussed in terms of methodological differences, and guidelines for future investigation of the law were suggested.

APPENDIX A

Statements of the Law

The original version of what has come to be known as the Yerkes-Dodson law appeared in an early article by Yerkes and Dodson (1908) which reported investigations of learning in the dancing mouse. Following a discussion of their results these investigators stated several conclusions, the last of which was:'

> As the difficultness of discrimination is increased the strength of that stimulus which is most favorable to habit formation approaches the threshold (Yerkes & Dodson, 1908, p. 482).

"Strength of stimulus" here meant "level of shock" and "threshold" referred to the threshold of stimulation. Yerkes and Dodson (1908) indicated that this conclusion applied to the particular experimental situation which was used, but suggested that other studies might be carried out to investigate the generality of this statement. On the basis of this research Yerkes (1909) later tenatively proposed the above conclusion as a "law of habit formation" to be a model for other laws in psychology, but he clearly delimited this formulation by explaining that it might not hold for other species, for other forms of motivation, or for tasks other than those involving a brightness discrimination.

In reporting a later experiment which attempted to extend the generalization to chickens, Cole (1911) restated the original conclusion to read:

> There appears to be an optimal strength of stimulus for each degree of difficulty of discrimination and the intensity of this optimal stimulus is less the more difficult the discrimination which is to be made (Cole, 1911, p. 111).

Cole referred to the observed effect of the shock level in the easy discrimination task as a law, since this effect was found in both this experiment and in the experiment by Yerkes and Dodson (1908), but since the results on other tasks were less conclusive, he did not label as a law the more general relationship discovered by Yerkes and Dodson (1908). Cole also apparently introduced the word "optimal" in referring to the shock level, for this wording is not found in the original version. A subsequent article by Dodson (1915) made reference to the findings of the original experiment, but there was no statement of the earlier conclusion and no implication that it should be construed as a law.

Under the topic of "Punishment" in his text on motivation, Young (1936) described the experiment by Yerkes and Dodson (1908) and then stated:

> There is thus an optimum intensity of punishment for a given degree of difficulty of the task, and if the strength of stimulation is increased beyond this optimum, the speed of learning is decreased rather than increased (Young, 1936, p. 283).

Young then quoted the original conclusion by Yerkes and Dodson and referred to it as the "Yerkes Dodson Law".

Broadhurst (1957) compressed the wording but expanded the meaning of the law when he wrote in the indroduction of his article:

> The Yerkes-Dodson Law which states that the optimum motivation for a learning task decreases with increasing difficulty has been shown to hold for several species (Broadhurst, 1957, p. 345).

Broadhurst concluded that his experiment had confirmed the law, apparently meaning this formulation of it.

In discussing the relation of the intensity of noxious stimulation to performance, Bindra (1959) mentioned that the relation is typically a U-shaped function and then stated:

> This type of relation between intensity of noxious stimulation and acquisition of habit strength was first suggested by Yerkes and Dodson (1908). These investigations show that the "amount" of punishment terminated affects the acquisition of an avoidance response (Bindra, 1959, p. 160).

Hall (1961) described the experiment by Yerkes and Dodson (1908) and quoted their conclusion, but also added another interpretation of it:

> Their (Yerkes and Dodson's) findings indicated an optimum intensity of a motivational antecedent for a given degree of difficulty of task; if intensity was increased beyond the optimum, the speed of learning decreased (Hall, 1961, p. 164).

Brown (1961) also contributed an interpretation which may be considered as a restatement since it closely parallels Broadhurst's (1957) formulation. In discussing the results of the Yerkes and Dodson (1908) experiment, Brown (1961) wrote:

> . . . on difficult problems performance was poorer with weak and with strong shocks than with shocks of intermediate strength. This latter finding led them to propose a general principle, since known as the Yerkes-Dodson law, to the effect that there is an optimal motivational level for learning, which tends to decrease as problem difficulty increases (Brown, 1961, p. 91).

Many other explanations and reformulations of the law are to be found in the literature. The above presentation is not intended to be an exhaustive survey, but merely a sampling of the various interpretations which have appeared. These examples clearly illustrate some of the changes which have been made and which contribute to the present confusion. The important point to note is that Yerkes' (1909) statement of the law was strictly qualified to certain experimental conditions, but later statements of the law by other writers have tended to ignore those qualifications while ascribing a greater, and perhaps unjustified, generality to the law.

APPENDIX B

Related Studies

The purpose of the original experiment by Yerkes and Dodson (1908) was to gain knowledge concerning the relation of motivation level to the rate of learning in the dancing mouse. The _Ss_ were required to pass through one of two openings surrounded by either white or black cardboard in order to escape to a nest-box. The white opening was designated as correct and the animal received an electric shock if it attempted to escape through the black opening. Three different levels of task difficulty were defined by the area of cardboard present which would reflect light. The shock level was varied by means of an inductorium and was measured in Martin units. There were 5 groups receiving different levels of shock in the easy task, 3 groups in the medium task, and 4 groups in the difficult task. Each group contained either 2 or 4 _Ss_. Each _S_ was given 10 trials per day and run to a criterion of 30 consecutive correct trials.

The results indicated that the difficult task was learned most rapidly at low levels of shock while the easy and medium-difficulty tasks were learned most rapidly at higher levels of shock. From these results the experimenters drew their now-famous conclusion which has evolved into a broad generalization known as the Yerkes-Dodson law.

It is important to remember that these early investigators did not have the benefit of parametric statistical methods in analyzing their data. The only analysis they were able to perform was to compute the mean scores of the various groups and plot these means on a graph. Subsequent inspection of data summarized in this way provided the basis for the conclusions drawn. Considering the limitations under which they worked, the early researchers did a commendable job.

Cole (1911) later performed an experiment which attempted to extend the generalization to chickens. The apparatus was essentially the same as that used by Yerkes and Dodson (1908), with the exception that milk-glass discrimination panels illuminated by lights were used as discriminanda instead of black and white cardboard. Cole defined three levels of task difficulty by means of visual inspection of various light combinations. Three levels of shock were employed, thus providing a 3 x 3 factorial experiment with 5 or 6 Ss in each treatment. The learning criterion in this instance was 20 consecutive correct trials. Many of the birds died during the course of the experiment and some never reached criterion, but by assigning very high scores to those Ss that did not learn, Cole was able to conclude that his results supported the conclusions drawn by Yerkes and Dodson (1908).

Dodson (1915) attempted to extend the generalization even further by using kittens as Ss in an experiment which was very similar to the original study (Yerkes & Dodson, 1908). There were fewer experimental groups, however, and most groups contained only 2 Ss. The results were consistent with the espoused law, and Dodson concluded that they were "in accord" with the original findings, but the severe limitations of the experimental conditions restrict any conclusions which might be drawn from this study. A later investigation by Dodson (1917) with rats as Ss employed four levels of shock and four levels of food deprivation.

Since only one task was used the results cannot provide verification of the law, but they are of interest since intermediate levels of motivation produced the most rapid learning with both forms of motivation. The most rapid learning with hunger motivation was found at 41 hours of food deprivation and the most rapid learning with shock motivation was found at a level of 75 Martin units, an intermediate level. In all cases, however, the shock groups showed faster learning than the hunger groups.

Broadhurst (1957) performed an experiment designed to increase the generality of the law by extending it to include the rat. A factorial design was used with 120 male albino rats as Ss. The apparatus was an underwater Y-maze with a discrimination panel in each arm which could be differentially illuminated. Three ratios of illumination between the panels were selected to serve as easy, moderate, and difficult tasks. The type of motivation employed was air deprivation which was varied by retaining Ss in an underwater startbox for different numbers of seconds. Following pretraining, during which the panels were unlit, all Ss were given 100 learning trials. The brighter panel was designated as correct and the correct side was randomly varied. The measure taken was correct trials out of 100 and an analysis of variance was performed on the scores obtained. A significant Task Difficulty effect and a significant Motivation x Task Difficulty interaction were found. Broadhurst interpreted these findings as confirming the Yerkes-Dodson law, but Brown (1965) has questioned this interpretation, claiming that the demonstration of a motivation effect within each task is also required.

There have been other studies in which both task difficulty and motivation have been varied, but most of these studies were not designed to test the Yerkes-Dodson law, and the results have been generally inconclusive. Hammes (1956) used 2 levels of task difficulty and 3 levels of shock motivation, but since only the difficult task was affected by motivation these results could not confirm the law. Miles (1959) found that performance was independent of drive level in squirrel monkeys when 3 levels of food deprivation and 2 levels of task difficulty were used. In a study with humans as Ss Chiles (1958) used difficult and easy paired associates with shock (high drive) and nonshock (low drive) groups. The high drive group did better on both the easy and difficult items, a result which is not inconsistent with the Yerkes-Dodson law, but which adds no new evidence for the law since the levels of the variables used were quite limited.

Studies of paired-associate learning in high- and low-anxious humans (Spence, 1956, 1958; Spence, Farber & McFann, 1956; Spence, Taylor & Ketchel, 1956) have often found an interaction between anxiety level and the type of items used, with high-anxious Ss superior to low-anxious Ss on easy lists and low-anxious Ss initially superior to high anxious Ss on difficult lists. If anxiety is considered a drive, or at least a correlate of drive, then these findings are consistent with the Yerkes-Dodson law and provide some supporting evidence for it. Other studies with humans have used high and low anxiety as drive levels with easy and difficult verbal and motor tasks, but while some of the results have been consistent with the law, other results have been conflicting (for reviews see Taylor, 1956; Jensen, 1958; Saranson, 1960; Hall, 1961; Brown, 1961.)

The Yerkes-Dodson law predicts that drive level will have a significant effect on the number of correct responses or number of errors during acquisition of a learning task. Experiments using several drive levels, but usually one task only, have yielded conflicting results. No effect of drive on error scores has been found in experiments with rats by Teel (1952), Armus (1958), Hillman, Hunter and Kimble (1953), and Eisman, Asimow and Maltzman (1956). However, experiments by Heyer (1951) and O'Kelly and Heyer (1951) have found a significant effect of drive on relearning errors and trials to criterion, and Birch (1955) found a U-shaped effect of drive on insightful problem solution in chimps. Any generalization from the data of these experiments would be difficult to make, however, due to the diversity of procedures and tasks employed.

Further research needs to be done in this area to clarify the relationship of task difficulty and drive level, and several writers (Broadhurst, 1957; Brown, J., 1961; Brown, W. P., 1965; Cofer & Appley, 1964) have suggested that the Yerkes-Dodson law should be tested in other situations using different tasks and different drives. The present experiment accomplished this in part, since food deprivation had not previously been the drive used in a test of the law.

APPENDIX C

Preliminary Study

A preliminary study was performed to isolate tasks of various levels of difficulty which could be used in the present experiment. The Ss were 50 male albino rats, approximately 90 days of age, obtained from Northwest Rodent Supply Co., Pullman, Washington. All Ss were placed on a 22-hr. deprivation schedule 2 days before the beginning of handling experience. Each S was then handled in the experimental room for 3 min. a day on 15 consecutive days with food available during handling. At the end of the handling period the Ss were randomly assigned to 5 task groups of 10 Ss each. The apparatus used was the discrimination box described earlier (p. 7) and the 5 filter combinations investigated (1 combination for each task group) were as follows: Task 1 - .5 and 3.5 lru; Task 2 - .5 and 2.5 lru.; Task 3 - 1.5 and 3.0 lru.; Task 4 - 2.0 and 3.0 lru.; Task 5 - 2.0 and 2.5 lru. In order to test for a brightness effect the darker stimulus was positive for half of the Ss in each task group and the brighter stimulus was positive for the remaining Ss. The procedure, reinforcement, and learning criterion were the same as that described in the above study (p. 9).

Mean trials to criterion are presented in Table A, and the results of an analysis of variance with a factorial design (5 task levels x 2 brightness levels) are presented in Table B. The Task effect was highly

Table A.

Mean trials to criterion by task and brightness groups in preliminary study.

		Subgroup Means	Task Mean
Task 1:	Bright	64.2	77.7
	Dark	91.2	
Task 2:	Bright	93.4	87.5
	Dark	81.6	
Task 3:	Bright	96.6	99.0
	Dark	101.4	
Task 4:	Bright	144.6	139.5
	Dark	134.4	
Task 5:	Bright	227.2	219.3
	Dark	211.4	

Table B

Summary of trials to criterion analysis of variance in preliminary study.

Source of Variance	Sum of Squares	df	Mean Square	F
Task (T)	134,214.8	4	33,553.7	41.88°
Brightness (B)	18.0	1	18.0	n.s.
T x B	3,094.4	4	773.6	n.s.
Error	32,046.8	40	801.17	
Total	169,374.0	49		

°$p < .005$

significant (*F* = 41.88, *df* = 4/40, *p* $< .005$), and further comparison of trials to criterion for each task group with Duncan's New Multiple Range Test and the Mann-Whitney *U* Test indicated that tasks 1, 4, and 5 differed significantly ($p < .05$) from one another. These three tasks, then, comprised the three levels of task difficulty in the present experiment. No significant brightness effect or Task x Brightness interaction was expected and none was found.

APPENDIX D

Supplementary Data

In the analysis of variance of trials to criterion in the present experiment, the F-ratio obtained for the interaction term was very close to the minimum required for statistical significance. There is, therefore, some reason to suspect that violating the assumptions underlying the analysis of variance may have influenced the significance of the interaction. To check on that possibility the data were analyzed in another way. The scores of the 4-hr. group were omitted since that group seemed to be contributing most to non-normality of the data and did not, in any case, provide an optimum drive level for any of the tasks investigated. The remaining data from the 10-, 22-, 36-, and 44-hr. groups were subjected to Bartlett's Test, which indicated no significant heterogeneity of variance ($X^2 = 16.805$, $df = 11$, $p > .10$). An analysis of variance performed without the 4-hr. groups (Table C) revealed a significant Task effect ($F = 37.70$, $df = 2/84$, $p < .005$), Drive effect ($F = 30.24$, $df = 3/84$, $p < .005$), and Task x Drive interaction ($F = 3.75$, $df = 6/84$, $p < .005$). Duncan's New Multiple Range Test and the Mann-Whitney U Test were again used to compare drive groups within each task, and the results from those comparisons were identical to the results from comparisons made with all groups present (Tables 3 & 4). The results of the analysis with the 4-hr. groups

Table C.

Summary of trials to criterion analysis of variance for 10-, 22-, 36-, and 44-hr. deprivation groups.

Source of Variance	Sum of Squares	df	Mean Square	F
Task (T)	93,657.00	2	46,828.50	37.70*
Drive (D)	112,669.11	3	37,556.37	30.24*
T x D	27,951.74	6	4,658.62	3.75*
Error	104,314.85	84	1,241.84	
Total	338,592.70	95		

*$p < .005$

omitted meet the preset requirements for a demonstration of the law and thus strengthen the conclusions drawn from the results of the analysis with all drive groups included.

The data were also analyzed in terms of total errors to criterion, but since an analysis of error scores was not originally planned, those results are included here for their value as supplementary information. The results of an analysis of variance of total error scores are presented in Table D, and the mean errors to criterion for all groups are presented graphically in Figure A. The Task effect ($F = 6.18$, $df = 2/105$, $p < .005$), Drive effect ($F = 15.04$, $df = 4/105$, $p < .005$) and Task x Drive interaction ($F = 2.54$, $df = 8/105$, $p < .05$) were all found to be significant. Bartlett's Test revealed significant heterogeneity of variance ($X^2 = 66.53$, $df = 14$, $p < .01$), even when the data from the 4-hr. groups was omitted ($X^2 = 39.62$, $df = 11$, $p < .01$).

Means and standard deviations of total error scores are presented in Table E. In this analysis the lowest mean for both the Difficult and Medium tasks was found at the 22-hr. deprivation level and the lowest mean for the Easy task was found at the 44-hr. deprivation level. Duncan's New Multiple Range Test (Table F) indicated that the lowest drive group means from each task difficulty level was significantly different ($p < .05$) from two or more other means in the case of the Easy and Medium tasks, and from one other mean in the case of the Difficult task. The Mann-Whitney U Test (Table G) showed the drive group with the lowest mean from each task to be significantly different ($p < .05$) from at least two other means on that task in all cases.

Table D.

Summary of errors to criterion analysis of variance.

Source of Variance	Sum of Squares	df	Mean Square	F
Task (T)	19,214.34	2	9,607.17	6.18*
Drive (D)	93,522.92	4	23,380.73	15.04*
T x D	31,685.84	8	3,960.73	2.54**
Error	163,184.90	105	1,554.14	
Total	307,608.00	119		

*$p < .005$
**$p < .05$

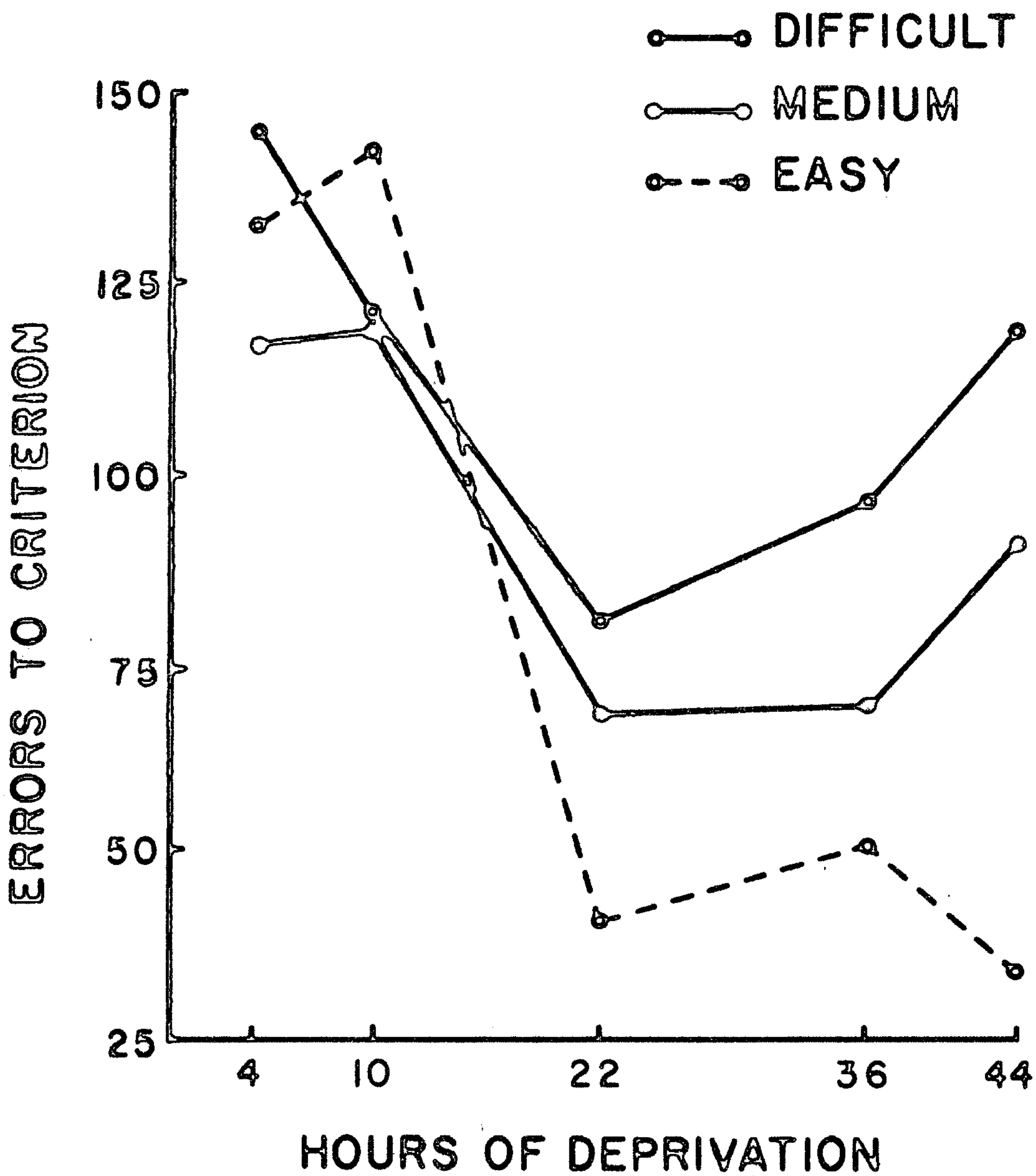

Figure A. Mean errors to criterion by level of task difficulty and drive level. Each point represents the mean total errors for 8 Ss.

Table E.

Means and standard deviations of total errors.

		Deprivation Condition				
		4 hr.	10 hr.	22 hr.	36 hr.	44 hr.
Easy Task	M	134.00	142.25	41.50	50.25	35.75
	SD	83.49	67.97	12.21	21.60	18.13
Medium Task	M	115.00	114.50	69.00	69.75	91.37
	SD	58.15	34.66	27.83	16.03	31.64
Difficult Task	M	144.62	117.5	78.75	96.87	119.12
	SD	44.68	25.71	17.68	15.35	30.89

Table F.

Significant differences between drive groups at each level of task difficulty in total errors (Duncan's New Multiple Range Test).

	Deprivation Condition			
	10 hr.	22 hr.	36 hr.	44 hr.
Easy Task				
4 hr.	-	.05	.05	.05
10 hr.		.05	.05	.05
22 hr.			-	-
36 hr.				-
Medium Task				
4 hr.	-	.05	.05	-
10 hr.		.05	.05	-
22 hr.			-	-
36 hr.				-
Difficult Task				
4 hr.	-	.05	.05	-
10 hr.		-	-	-
22 hr.			-	-
36 hr.				-

Table G.

Significance of differences in total errors between drive group with lowest mean and remaining drive groups at each level of task difficulty (Mann-Whitney U Test).

	Deprivation Condition				
	4 hr.	10 hr.	22 hr.	36 hr.	44 hr.
Drive Group with Lowest Mean					
Easy Task (44 hr., M = 35.75)	.002	.001	.139	.065	-
Medium Task (22 hr., M = 69.00)	.041	.010	-	.360	.097
Difficult Task (22 hr., M = 78.88)	.001	.001	-	.032	.002

These results are consistent with the results obtained from the analysis of trials to criterion, but are not identical to those results. The error score analysis indicated that the optimum drive level for both the Medium and Difficult tasks was 22 hr. deprivation while the analysis of trials to criterion indicated that the optimum level for the Medium task was 36 hr. deprivation and the optimum level for the Difficult task was 22 hr. deprivation. Since both measures are in agreement on the lowest means for the Difficult task (22 hr. deprivation) and the Easy task (44 hr. deprivation), the discrepancy between the two measures in determining the lowest mean for the Medium task may indicate that the optimum level for that task lies somewhere between the levels dictated by these two measures. That is, the optimum level for the medium task probably lies between 22 and 36 hrs. deprivation. If this is the case the conclusions already drawn are not discredited and the discrepancy only indicated that the levels of deprivation used were too widely separated to provide a more accurate estimation of the optimum level.

Running times for each animal were recorded manually with a Hunter Clock-Timer for the first 50 trials of the present experiment. Running times were transformed to reciprocals of medians for blocks of 5 trials and mean running scores were then calculated for each level of deprivation with task groups at each level combined. Mean run speeds are presented graphically in Figure B. An analysis of variance (Table H) of run speeds before task groups were combined indicated a significant Drive effect (*F* = 36.43, *df* = 4/105, *p* $<$.01) but no significant Task effect (*F* = 1.96, *df* = 2/105, *p* $>$.05) or Task x Drive interaction

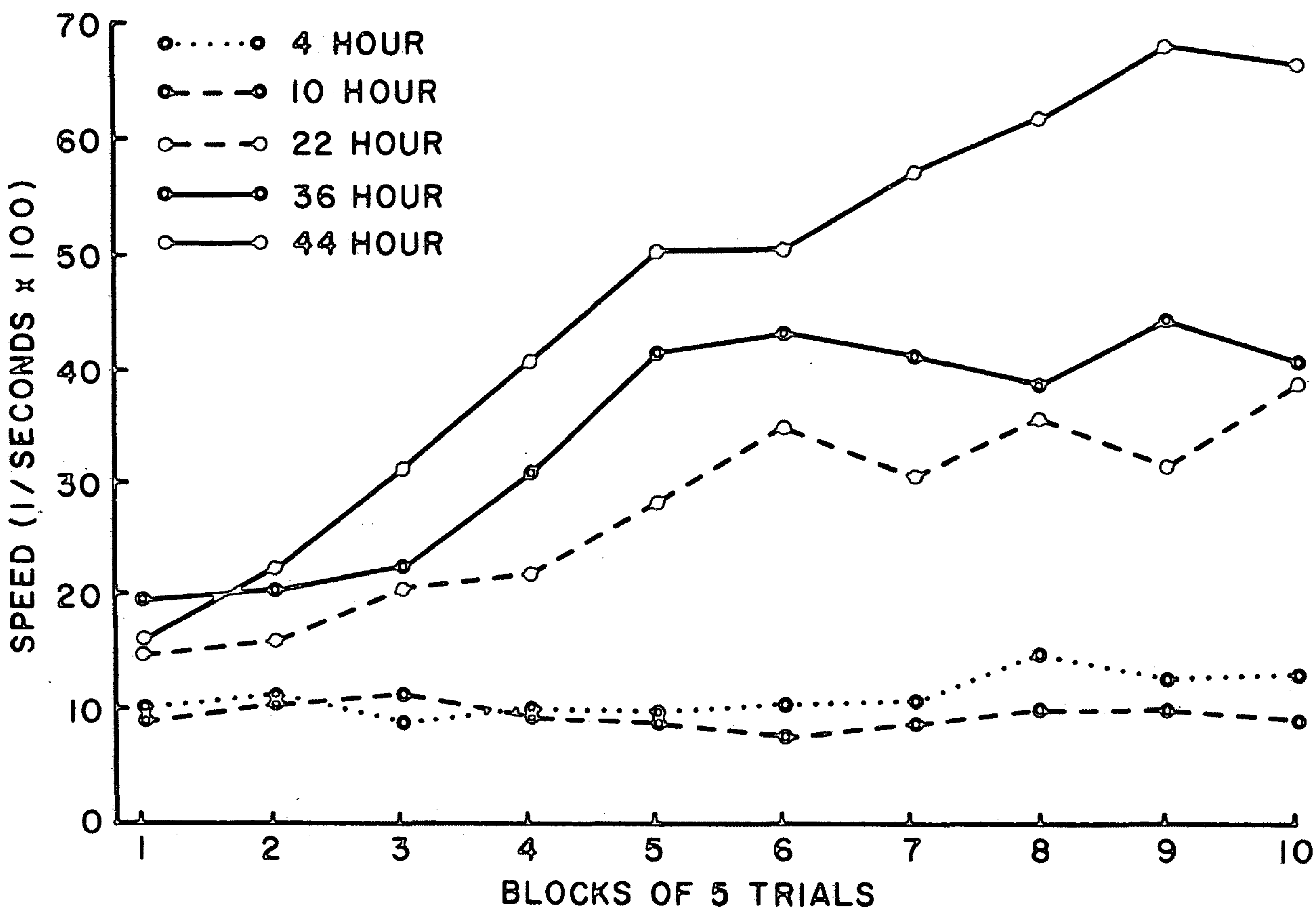

Figure B. Running speed as a function of trials.
Each point represents the mean of 24 Ss.

Table H.

Summary of run speed analysis of variance.

Source of Variance	Sum of Squares	df	Mean Square	F
Task (T)	140,940.0	2	70,470.0	1.96*
Drive (D)	5,222,227.0	4	1,305,556.75	36.43**
T x D	239,366.0	8	29,920.75	.834*
Error	3,762,516.0	105	35,833.49	
Total	9,365,050.0	119		

* $p > .05$
**$p < .01$

($\underline{F}$ = .834, $\underline{df}$ = 8/105, $\underline{p}$ > .05). The mean run speeds from each drive level on trials 35-50 were compared by means of the Mann-Whitney $\underline{U}$ Test (Table I) and all were found to be significantly different ($\underline{p}$ < .05) from one another. Duncan's New Multiple Range Test (Table J) was also used to compare the means on trials 35-50, and indicated that the 4- and 10-hr. deprivation groups were not significantly different from one another, but all other comparisons were significant ($\underline{p}$ < .05). Under the assumption that response strength is to some degree a behavioral index of drive strength, these results suggest that the drive levels were effectively manipulated in this experiment, with the exception that there was no differential manipulation of drive levels between the 4- and 10-hr. deprivation conditions.

Table I.

Significance of differences in run speeds between drive groups (Mann-Whitney U Test).

	Deprivation Condition			
	10 hr.	22 hr.	36 hr.	44 hr.
4 hr.	.05	.05	.05	.05
10 hr.		.05	.05	.05
22 hr.			.05	.05
36 hr.				.05

Table J.

Significance of differences in run speeds between drive groups (Duncan's New Multiple Range Test).

	Deprivation Condition			
	10 hr.	22 hr.	36 hr.	44 hr.
4 hr.	-	.05	.05	.05
10 hr.		.05	.05	.05
22 hr.			.05	.05
36 hr.				.05

REFERENCES

Armus, H. L. Drive level and habit reversal. Psychological Reports, 1958, 4, 31-34.

Bindra, D. Motivation: A systematic reinterpretation. New York: Ronald Press, 1959.

Birch, H. G. Motivation in insightful problem solving. In McClelland, D. C. Studies in Motivation. New York: Appleton-Century-Crofts, 1955.

Boneau, C. A. The effects of violations of assumptions underlying the t test. Psychological Bulletin, 1960, 57, 49-64.

Box, G. E. P. Non-normality and tests on variances. Biometrika, 1953, 40, 318-335.

Broadhurst, P. L. Emotionality and the Yerkes-Dodson Law. Journal of Experimental Psychology, 1957, 54, 345-352.

Broadhurst, P. L. The interaction of task difficulty and motivation: the Yerkes-Dodson Law revived. Acta Psychologica, 1959, 16, 321-338.

Brown, J. The motivation of behavior. New York: McGraw-Hill, 1961.

Brown, W. P. The Yerkes-Dodson Law repealed. Psychological Reports, 1965, 17, 663-666.

Chiles, W. D. Effects of shock-induced stress on verbal performance. Journal of Experimental Psychology, 1958, 56, 159-165.

Cofer, C. N., & Appley, M. H. Motivation: Theory and research. New York: Wiley & Sons, 1964.

Cole, L. W. The relation of strength of stimulation to rate of learning in the chick. Journal of Animal Behavior, 1911, 1, 111-124.

Dodson, J. D. The relation of strength of stimulus to rapidity of habit-formation in the kitten. Journal of Animal Behavior, 1915, 5, 330-336.

Dodson, J. D. Relative values of reward and punishment in habit-formation. Psychobiology, 1917, 1, 231-276.